A Probabilistic Analysis of Some Tree Algorithms

Hanene Mohamed, Philippe Robert

A PROBABILISTIC ANALYSIS OF SOME TREE ALGORITHMS

HANÈNE MOHAMED AND PHILIPPE ROBERT

Abstract. In this paper a general class of tree algorithms is analyzed. It is shown that, by using an appropriate probabilistic representation of the quantities of interest, the asymptotic behavior of these algorithms can be obtained quite easily without resorting to complex analysis techniques as it is usually the case. This approach gives a unified probabilistic treatment of these questions. It simplifies and extends some of the results known in this domain.

Contents

1. Introduction

A splitting algorithm is a procedure that divides recursively into subgroups an initial group of n items until each of the subgroups obtained has a cardinality strictly less than some fixed number D. These algorithms have a wide range of applications.

— Data structures. These are algorithms on data structures used to sort and search. They are sometimes referred to as divide and conquer algorithms. See Cormen *et al.* [1] and Knuth [22] for a general presentation and Flajolet and Sedgewick [38], Mahmoud [27] and Szpankowski [39] for their analysis with analytical methods.

— Communication Networks. These algorithms are used to give a distributed access to a common communication channel that can transmit only one message per time unit. See Capetanakis [4], Tsybakov and Mikhaĭlov [40] and Ephremides and Hajek [9].

Date: December 9, 2004.

Key words and phrases. Splitting Algorithms. Divide and Conquer Algorithms. Unusual Laws of Large Numbers. Asymptotic Oscillating Behavior. Data Structures. Tries. Renewal Theorem.

— Distributed systems. Some algorithms use a splitting technique to select a subset of a group of identical communicating components. See Janson and Szpankowski [17] and Raz *et al.* [35].
— Statistical tests. A test, performed on a group of individuals, indicates if at least one of these individuals has some characteristics (like a disease if this is blood testing). The purpose is to minimize the number of tests to identify individuals with the specified characteristic as quickly as possible. See Wolf [42]

Formally, a splitting algorithm can be described as follows:

Splitting Algorithm $\mathcal{S}(n)$

— Termination Condition.
If $n < D$ $\longrightarrow$ Stop.
— Tree Structure.
If $n \geq D$, randomly divide n into $n_1, \ldots, n_G$, with $n_1 + \cdots + n_G = n$ where G is a random variable with some fixed distribution.
$\longrightarrow$ Apply $\mathcal{S}(n_1)$, $\mathcal{S}(n_2)$, …, $\mathcal{S}(n_G)$.

1.1. **Description.** The algorithm starts with a group of size n items. This set is randomly split into G subgroups, the distribution of G is given by $\mathbb{P}(G = \ell) = p_\ell$, where (p_ℓ) is a probability distribution on $\{2, 3, \ldots\}$. Now, conditionally on the event $\{G = \ell\}$, for $1 \leq i \leq \ell$, an item is sent into the ith subgroup with probability $V_{i,\ell}$, where $V_\ell = (V_{i,\ell}; 1 \leq i \leq \ell)$ is a random probability vector on $\{1, \ldots, \ell\}$. It can also be seen as a vector of random weights on the ℓ arcs of the branching procedure on which each of the n items perform a random walk.

If N_i is the cardinality of the ith subgroup then, conditionally on the event $\{G = \ell\}$ and on the random variables $V_{1,\ell}, V_{2,\ell}, \ldots, V_{\ell,\ell}$, the distribution of the vector $(N_1, \ldots, N_\ell)$ is multinomial with parameter n and $(V_{1,\ell}, V_{2,\ell}, \ldots, V_{\ell,\ell})$,

$$\mathbb{P}\left((N_1, \ldots, N_\ell) = (m_1, \ldots, m_\ell)\right) = \frac{n!}{m_1!\, m_2! \cdots m_\ell!} \prod_{k=1}^{\ell} (V_{k,\ell})^{m_k},$$

for $(m_i) \in \mathbb{N}^n$ such that $m_1 + \cdots + m_\ell = n$. If the ith subgroup, $1 \leq i \leq n$, is such that $N_i < D$, the algorithm stops for this subgroup. Otherwise, it is applied to the ith subgroup: a variable G_i, with the same distribution as G, is drawn and this ith subgroup is split into G_i subgroups, and so on …

Such a random splitting has been introduced by Devroye [7] where the asymptotic expansion of the depth of the associated tree is investigated.

Examples.

— Knuth's Algorithm. When $\mathbb{P}(G = 2) = 1$, $D = 2$ and $V_{1,2} \equiv V_{2,2} \equiv 1/2$, this is one of the oldest algorithms of this kind. It has been analyzed by Knuth in 1973.
— Symmetrical Splitting Algorithm. This is the case where $V_{i,n} \equiv 1/n$ for any $n \geq 2$ and $1 \leq i \leq n$.
— Q-ary algorithm. If $\mathbb{P}(G = Q) = 1$ and $D = 2$ this is the Q-ary resolution algorithm with blocked arrivals analyzed by Mathys and Flajolet [29].

See also Devroye [7] for other examples. Quite naturally, such an algorithm can be graphically represented with a tree as shown by Figure 1.

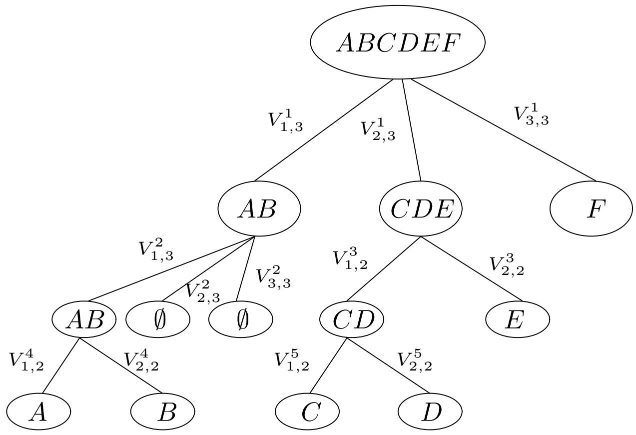

FIGURE 1. Splitting Algorithm with $D = 2$, two sets of *random* weights $(V_{1,2}, V_{2,2})$ and $(V_{1,3}, V_{2,3}, V_{1,3})$, G a random variable with values in $\{2, 3\}$ and the initial items are A, B, C, D, E and F.

Splitting Measure. As it will be seen in the following, the key characteristic of this splitting algorithm is a probability distribution $\mathcal{W}$ on $[0, 1]$ defined with the branching distribution (the variable G) and the weights on each arc (the vector $(V_{1,G}, \dots, V_{G,G})$). The asymptotic behavior of the algorithm is expressed naturally in terms of the distribution $\mathcal{W}$.

Definition 1. *The* splitting measure *is the probability distribution* $\mathcal{W}$ *on* $[0, 1]$ *defined by, for a non-negative Borelian function* f*,*

$$\int f(x)\, \mathcal{W}(dx) = \mathbb{E}\left(\sum_{i=1}^{G} V_{i,G} f(V_{i,G})\right) = \sum_{\ell=2}^{+\infty}\sum_{i=1}^{\ell} \mathbb{P}(G=\ell)\mathbb{E}(V_{i,\ell} f(V_{i,\ell})). \tag{1}$$

Throughout the paper, it is assumed that, almost surely $G \geq 2$, and that there exists some $\delta > 0$ such that the relation

$$\sup_{\ell \geq 2} \sup_{1 \leq i \leq \ell} V_{i,\ell} \leq \delta < 1 \tag{A}$$

holds almost surely, in particular $\mathcal{W}([0, \delta]) = 1$. These conditions imply in particular the non-degeneracy of the splitting mechanism.

Definition 2. *A splitting measure* $\mathcal{W}$ *is* exponentially arithmetic *if, there exists some* $\lambda > 0$ *such that*

$$\mathcal{W}\left(\left\{e^{-n\lambda} : n \geq 1\right\}\right) = 1,$$

the largest λ *satisfying this relation is defined as the* exponential span *of* $\mathcal{W}$*.*

If A is some random variable with distribution $\mathcal{W}$, then $\mathcal{W}$ is exponentially arithmetic with exponential span λ if and only if the distribution of $-\log(A)$ is arithmetic with span λ. See Section 5.

Examples.

— Knuth's Algorithm, $\mathbb{P}(G=2)=1$, $D=2$ and $V_{1,2}\equiv V_{2,2}\equiv 1/2$.
In this case
$$\mathcal{W}(dx)=\delta_{1/2}$$
where δ_x is the Dirac distribution at x, $\mathcal{W}$ is exponentially arithmetic with exponential span $\log 2$.

— Symmetrical Splitting Algorithm.
$$\mathcal{W}(dx)=\sum_{n\geq 2} P(G=n)\delta_{1/n},$$
the exponential span is $\log D$ where D is the largest integer p such that the support of the random variable G is contained in $p\mathbb{N}$.

— Q-ary algorithm, $\mathbb{P}(G=Q)=1$, $D=2$, $V_{1,Q}=p_1,\ldots, V_{Q,Q}=p_Q$.
$$\mathcal{W}(dx)=p_1\delta_{p_1}+p_2\delta_{p_2}+\cdots+p_Q\delta_{p_Q},$$
the distribution $\mathcal{W}$ is exponentially arithmetic if and only if all the real numbers $\log p_i/\log p_j$, $1\leq i<j\leq Q$, are rational.

The Cost of a Splitting Algorithm. For such an algorithm, an important quantity is the number of operations required until the algorithm stops, i.e. when all the subgroups have a cardinality less or equal to D. Denote by R_n this quantity when the number of initial items is n, then clearly

— $R_n=1$ when $n<D$;

— For $n\geq D$,
$$R_n \stackrel{\text{dist.}}{=} 1+R_{1,N_1^n}+\cdots+R_{G,N_G^n}, \tag{2}$$
where conditionally on the event $\{G=\ell\}$ and the random variables $V_{1,\ell}$, $V_{2,\ell}$, …, $V_{\ell,\ell}$,
(1) The vector $(N_1^n,\ldots,N_\ell^n)$ has a multinomial distribution with parameter n and $(V_{1,\ell},V_{2,\ell},\ldots,V_{\ell,\ell})$;
(2) For $(p_i)\in\mathbb{N}^\ell$, the variables $R_{1,p_1},\ldots,R_{\ell,p_\ell}$ are independent;
(3) For $1\leq i\leq \ell$, the variable R_{i,p_i} has the same distribution as R_{p_i}.

The variable R_n is simply the number of nodes of the associated tree, see Figure 1.

1.2. **Unusual Laws of Large Numbers.** Note that, since the splitting procedure is random, the variable R_n is a *random variable*. With the language of communication networks, this quantity can be thought as the total time to transmit n initial messages. If $\mathbb{E}(R_n)$ is its expected value, $\mathbb{E}(R_n)/n$ is the average transmission time of one message among n. From a probabilistic point of view, it is natural to expect that the sequence (R_n) satisfies a kind of law of large numbers, i.e. that $(\mathbb{E}(R_n)/n)$ converges to some quantity α. The constant α is, in some sense, the asymptotic average transmission time of a message. Curiously, this law of large numbers does not always hold. In some situations, the sequence $(\mathbb{E}(R_n)/n)$ does not converge at all and, moreover, exhibits an oscillating behavior.

When the splitting degree is constant and equal to 2 and $V_{1,2}\equiv V_{2,2}\equiv 1/2$ (the items are equally divided among the two subgroups), these phenomena are quite well known. They have been analyzed using complex analysis techniques, functional transforms (and their associated inversion procedures) by Knuth [22], Flajolet *et al.* [12], Louchard and Prodinger [26] and many others. See Hofri [15], Mahmoud [27] and Flajolet and Sedgewick [38] for a comprehensive treatment of this approach. See also Devroye [8] for a survey of the domain. Robert [36] proposed an alternative, elementary, method to get the asymptotic behavior of some related oscillating sequences without using complex analysis.

When the splitting degree is constant and equal to Q but the items are *not* equally divided among the subgroups, studies are quite rare. Using complex analysis techniques, Fayolle *et al.* [11] obtains the asymptotic behavior of the associated sequence $(\mathbb{E}(R_n))$. Mathys and Flajolet [29] gives a sketch of a generalization of this study when Q is arbitrary.

Some alternative approaches.

- Some laws of large numbers have been proved by Devroye [8] in a quite general framework for various functionals of the associated trees. Talagrand's concentration inequalities are the main tools in this study. In our case, it would consist in proving that, the distribution of the random variable $R_n/\mathbb{E}(R_n)$ is sharply (with an exponential decay) concentrated around 1. Results on limiting distributions such as, central limit theorems, do not seem to be accessible with this method.
- Clement *et al.* [6] analyzed related algorithms in the more general context of dynamical systems. By using an Hilbertian setting, they show that the first order behavior of the algorithms is expressed in terms of the spectrum of a functional operator, the transfer operator. Getting explicit results in this way requires therefore a good knowledge of some eigenvalues of the transfer operator.

A dynamic version of this class of algorithms is investigated in Mohamed and Robert [32]. The splitting procedure is the same but, in the language of branching processes, an immigration occurs at every leave of the associated tree, i.e. new items arrive every time unit. This dynamic feature complicates the problem. In this case, an additional probabilistic tool has to be used: an autoregressive process with moving average plays an important role.

1.3. **Related Problems.**

Fragmentation Processes.
A continuous version of a splitting algorithm could be defined as follows: an initial mass of size x is randomly split into several pieces and, at their turn, each of the pieces is randomly split, ... A class of such models has been recently investigated. The fragmentation of each mass occurs after some independent exponential time with a parameter depending, possibly, of its mass. See Bertoin [3], Miermont [31] and references therein. The problems considered are somewhat different: regularity properties of associated Markov processes, duality, rate of decay of individual masses, loss of mass, asymptotic distributions, ... A splitting algorithm is just a recursive fragmentation of an integer into integer pieces until each of the components has a size less than D. In a continuous setting, an analogue of the algorithms considered here would consist in stopping the fragmentation process of a mass as soon as its value is below some threshold $\varepsilon > 0$.

Random recursive decompositions.
As it will be (easily) seen, a splitting algorithm can also be described as a random recursive splitting of the interval $[0, 1]$. For example in the case of a dyadic splitting,

starting from the interval $[0, 1]$, two subintervals I_1, I_2 are created and each of them is split at its turn and so on ...

These random recursive decompositions have been considered from the point of view of the geometry of the boundary points by various authors, to express the Hausdorff dimension of this set of points in particular. Mauldin and Williams [30], Waymire and Williams [41] consider decompositions of the interval $[0, 1]$ which are not necessarily conservative, i.e. when $|I_1|+|I_2| < 1$ holds with positive probability in the dyadic case.

Hambly and Lapidus [14] or Falconer [10] consider decompositions of the interval $[0, 1]$ from the point of view of the lengths of the associated subintervals. The interval $[0, 1]$ is represented by a non-increasing sequence (L_n) whose sum is 1. For $n \geq 1$, L_n is the length of the nth largest interval of the decomposition. This description is similar to the classical representation of fragmentation processes. See Pitman [33].

In this setting, multiplicative cascades and martingales introduced by Mandelbrot [28] and Kahane and Perrière [18] show up quite naturally. They have been analyzed quite extensively, see Liu [24], Barral [2] and references therein.

1.4. **An Overview.** The purpose of this paper is twofold. First, it considers splitting algorithms with a random (and possibly unbounded) degree of splitting generalizing the previous studies in this domain. Secondly, and this is in fact the main point of the paper, it proposes a probabilistic approach that simplifies much the analysis of these algorithms. Moreover, as a byproduct, a new direct representation of the asymptotic oscillating behavior is established.

The analysis proposed in this paper also starts from Equation (2), but its treatment is significantly different from the analytic approach. After some transformation, Equation (2) is interpreted as a probabilist equation which is iterated by using appropriate independent random variables. Following the method of Robert [36], the next step is to perform a probabilistic de-Poissonnization and, by using Fubini's Theorem conveniently, to represent the quantity $\mathbb{E}(R_n)$ by using a Poisson point process on the real line. The final, crucial, step which differs from Robert [36], consists in using the key renewal Theorem to get the asymptotic behavior of the sequence $(\mathbb{E}(R_n))$.

The approach is elementary, its main advantage on the analytic treatment lies certainly in the use of the renewal Theorem which gives *directly* the asymptotic behaviour.

Results of the paper. Section 2 gives a useful representation for the average cost of the algorithm. The main result of the paper for the asymptotic cost is the following theorem in Section 3. This is a summary of Propositions 9 and 11.

Theorem 3 (Asymptotics of the Average Cost)**.** *For a splitting algorithm, under the condition*

$$\int_0^1 \frac{|\log(y)|}{y} \mathcal{W}(dy) < +\infty, \tag{3}$$

— *if the splitting measure $\mathcal{W}$ is not exponentially arithmetic, then*

$$\lim_{n\to+\infty} \frac{\mathbb{E}(R_n)}{n} = \frac{\mathbb{E}(G)}{(D-1)\int_0^1 |\log(y)|\,\mathcal{W}(dy)}. \tag{4}$$

— *If the splitting measure $\mathcal{W}$ is exponentially arithmetic with exponential span $\lambda > 0$, as n gets large, the equivalence*

$$\frac{\mathbb{E}(R_n)}{n} \sim F\left(\frac{\log n}{\lambda}\right) \tag{5}$$

holds, where F is the periodic function with period 1 defined by, for $x \geq 0$,

$$F(x) = \frac{\mathbb{E}(G)}{\int_0^1 |\log(y)|\,\mathcal{W}(dy)} \frac{\lambda}{1-e^{-\lambda}} \int_0^{+\infty} \exp\left(-\lambda\left\{x - \frac{\log y}{\lambda}\right\}\right) \frac{y^{D-2}}{(D-1)!} e^{-y}\,dy$$

and $\{z\} = z - \lfloor z \rfloor$ is the fractional part of $z \in \mathbb{R}$.

Condition (3) is not really restrictive since the variable G is bounded in practice. This theorem covers and extends some of the results in this domain: for Knuth's Algorithm, Knuth [22] and for Q-ary algorithms with blocked arrivals, Mathys and Flajolet [29], see Corollaries 10 and 12.

Furthermore, when there are asymptotic periodic oscillations, the periodic function F involved is expressed directly and not in terms of its Fourier coefficients as it is usually the case. The expression of F generalizes the representation of Robert [36] obtained for Knuth's Algorithm.

The *distribution* of the sequence (R_n) (and not only its average) is investigated in Section 4. For simplicity, only the case where the variable G is constant and the variables $V_{\cdot,G}$ are equal to $1/G$ is considered. The purpose of this section is to show that the distribution of the Poisson transform of the sequence and more generally, the distribution of most of the functionals of the associated tree, can be expressed quite simply in terms of Poisson processes and uniformly distributed random variables.

Two representations of the distribution of the Poisson transform as a functional of Poisson processes are derived. As a consequence, a law of large numbers is proved when the number of initial items has a Poisson distribution (Poisson transform). Moreover, the asymptotic oscillating behavior of the algorithm is proved as a consequence of a *standard* law of large numbers. These unusual laws of large numbers are, in the end, in the realm of classical laws of large numbers.

The central limit theorem is also proved with a similar method in this case. This is a classical result, see Mahmoud [27], it is usually proved with complex analysis methods via quite technical estimations. It is proved here as a consequence of the *standard* central limit theorem for independent random variables. At the same time, a new representation of the asymptotic variance is obtained.

Section 5 recalls briefly some results and definitions concerning the renewal theorem.

2. General Properties

Throughout this paper, $(\mathcal{N}([0,x]))$ denotes a Poisson process with intensity 1, equivalently it can also be described as a non-decreasing sequence (t_n) such that $(t_{n+1} - t_n)$ is a sequence of i.i.d.random variables exponentially distributed with parameter 1. For $x \geq 0$, the variable $\mathcal{N}([0,x])$ is simply the number of t_n's in the interval $[0,x]$. See Kingman [19] for basic results on Poisson processes.

Equation (2) and the boundary conditions for the sequence (R_n) are summarized in the following relation, for $n \geq 0$,

$$R_n \stackrel{\text{dist.}}{=} 1 + R_{1,N_1^n} + \cdots + R_{G,N_G^n} - G\,\mathbb{1}_{\{n<D\}}.$$

therefore,

$$R_n - 1 \stackrel{\text{dist.}}{=} \sum_{i=1}^{G} \left(R_{i,N_i^n} - 1\right) + G\,\mathbb{1}_{\{n\geq D\}} \tag{6}$$

Definition 4. *The* Poisson transform *of a non-negative sequence (a_n) is defined as*

$$\sum_{n\geq 0} a_n \frac{x^n}{n!}\, e^{-x} = \mathbb{E}\left(a_{\mathcal{N}([0,x])}\right). \tag{7}$$

The following proposition gives useful representations of the Poisson Transform of the sequence of $(\mathbb{E}(R_n))$.

Proposition 5 (Poisson Transform of the sequence (R_n))**.** *For $x > 0$,*

$$\mathbb{E}\left(R_{\mathcal{N}([0,x])}\right) = 1 + \mathbb{E}(G)\mathbb{E}\left(\sum_{i=0}^{+\infty} \frac{1}{\prod_{k=1}^{i} W_k} \mathbb{1}_{\left\{t_D \leq x \prod_{k=1}^{i} W_k\right\}}\right). \tag{8}$$

where (W_i) is an i.i.d. sequence of random variables with distribution $\mathcal{W}$.

Proof. If n is a Poisson random variable with parameter x, the splitting property of Poisson variables (see Kingman [19] for example) shows that, conditionally on the event $\{G = \ell\}$ and on the variables $V_{1,\ell}, \ldots, V_{\ell,\ell}$, the variables N_i^n, $1 \leq i \leq l$ are independent and N_i^n has a Poisson distribution with parameter $xV_{i,l}$. Consequently, for $x > 0$, if

$$\Phi(x) \stackrel{\text{def}}{=} \frac{\mathbb{E}\left(R_{N([0,x])}\right) - 1}{x\mathbb{E}(G)}, \tag{9}$$

it is easily checked that $\mathbb{E}(G)\Phi(x) \to R_1 - R_0 = 0$ as $x \searrow 0$.

Since $\{N([0,x]) \geq D\} = \{t_D \leq x\}$, Equation (6) gives the relation

$$\Phi(x) = \sum_{\ell=2}^{+\infty} \mathbb{P}(G = \ell)\, \mathbb{E}\left(\sum_{i=1}^{\ell} V_{i,\ell}\Phi(xV_{i,\ell})\right) + \frac{1}{x}\mathbb{P}(t_D \leq x). \tag{10}$$

Equation (10) can then be rewritten as

$$\Phi(x) = \mathbb{E}(\Phi(xW_1)) + \mathbb{E}\left(\frac{1}{x}\mathbb{1}_{\{t_D\leq x\}}\right). \tag{11}$$

The iteration of Equation (11) shows that, for $n \geq 1$,

$$\Phi(x) = \mathbb{E}\left(\Phi\left(x\prod_{k=1}^{n} W_k\right)\right) + \mathbb{E}\left(\sum_{i=0}^{n-1} \frac{1}{x\prod_{k=1}^{i} W_k} \mathbb{1}_{\left\{t_D \leq x \prod_{k=1}^{i} W_k\right\}}\right).$$

The assumption on the variable G and the sequence of vectors (V_n) implies that, almost surely, the sequence $(\prod_{k=1}^{n} W_k)$ converges to 0. The function Φ can thus be represented as

$$\Phi(x) = \mathbb{E}\left(\sum_{i=0}^{+\infty} \frac{1}{x\prod_{k=1}^{i} W_k} \mathbb{1}_{\left\{t_D \leq x \prod_{k=1}^{i} W_k\right\}}\right).$$

The proposition has been proved. $\square$

From now on, throughout the paper, (W_i) will denote an i.i.d. sequence of random variables on $[0,1]$ with distribution $\mathcal{W}$.

Proposition 6 (Probabilistic de-Poissonnization). *For $n \geq D$, then*

$$\mathbb{E}(R_n) = 1 + \mathbb{E}(G)\mathbb{E}\left(\sum_{i=0}^{T\left(U_{(n)}^D\right)-1} \frac{1}{\prod_{k=1}^i W_k},\right) \tag{12}$$

where, for $0 < y < 1$,

$$T(y) = \inf\left\{i \geq 1 : \prod_{k=1}^i W_k < y\right\}$$

and $U_{(n)}^D$ is the Dth smallest variable of n independent, uniformly distributed random variables on $[0,1]$ independent of (W_i).

Proof. For $x > 0$, by decomposing with respect to the number of points of the Poisson process $(\mathcal{N}(t))$ in the interval $[0,x]$, one gets, for $0 < \alpha \leq 1$,

$$\begin{aligned}\mathbb{P}(t_D \leq x\alpha) &= \sum_{n=D}^{+\infty} \mathbb{P}\left(t_D \leq x\alpha, \mathcal{N}([0,x]) = n\right) \\ &= \sum_{n=D}^{+\infty} \mathbb{P}\left(t_D \leq x\alpha \mid \mathcal{N}([0,x]) = n\right)\mathbb{P}(\mathcal{N}([0,x]) = n).\end{aligned}$$

For $n \geq D$, conditionally on the event $\{\mathcal{N}([0,x]) = n\}$, the variable t_D has the same distribution as the D smallest random variable of n uniformly distributed random variables on $[0,x]$. When $x = 1$, denote by $U_{(n)}^D$ a variable with this conditional distribution. Clearly, by homogeneity, the variable $(t_D \mid \mathcal{N}([0,x]) = n)$ has the same distribution as $xU_{(n)}^D$. Finally, one gets the identity

$$\mathbb{P}(t_D \leq x\alpha) = \sum_{n=D}^{+\infty} \mathbb{P}\left(U_{(n)}^D \leq \alpha\right) \frac{x^n}{n!} e^{-x} = \mathbb{E}\left(\sum_{n=D}^{+\infty} \mathbb{1}_{\{U_{(n)}^D \leq \alpha\}} \frac{x^n}{n!} e^{-x}\right).$$

By using the independence of the sequence (W_i) and t_D in Equation (8), the last identity gives the relation

$$\mathbb{E}\left(R_{\mathcal{N}([0,x])}\right) = 1 + \mathbb{E}(G)\mathbb{E}\left(\sum_{i=0}^{+\infty} \frac{1}{\prod_{k=1}^i W_k} \sum_{n=D}^{+\infty} \mathbb{1}_{\{U_{(n)}^D \leq \prod_{k=1}^i W_k\}} \frac{x^n}{n!} e^{-x}\right).$$

By Fubini's Theorem and writing $1 = \exp(x)\exp(-x)$, this expression can be rewritten as

$$\sum_{n=0}^{D-1} \frac{x^n}{n!} e^{-x} + \sum_{n=D}^{+\infty} \left(1 + \mathbb{E}(G)\mathbb{E}\left(\sum_{i=0}^{+\infty} \frac{1}{\prod_{k=1}^i W_k} \mathbb{1}_{\{U_{(n)}^D \leq \prod_{k=1}^i W_k\}}\right)\right) \frac{x^n}{n!} e^{-x}.$$

The identification of Representation (7) of $\mathbb{E}\left(R_{\mathcal{N}([0,x])}\right)$ and the last identity gives Formula (12) $\square$

Corollary 7 (Symmetrical Q-ary Algorithm)**.** *When* $\mathbb{P}(G = Q) = 1$ *holds and* $V_{i,Q} \equiv 1/Q$*, for* $i = 1, \ldots Q$*, then for* $n \geq D$*,*

$$\mathbb{E}(R_n) = 1 + \frac{Q}{Q-1}\left(\mathbb{E}\left(Q^{\lceil -\log_Q U_{(n)}^D \rceil}\right) - 1\right) \tag{13}$$

with, for $0 \leq x \leq 1$*,*

$$\mathbb{P}\left(U_{(n)}^D > x\right) = \sum_{k=0}^{D-1} \binom{n}{k} x^k (1-x)^{n-k}.$$

From Equation (13), by using the fact that $nU_{(n)}^D$ converges in distribution as n tends to infinity, it is not difficult to get the asymptotic behavior of $\mathbb{E}(R_n)$. The general case, Equation (12), is slightly more complicated. One has to study the asymptotics of the series inside the expectation.

A Functional Integral Equation. If $R(x) = \mathbb{E}(R_{\mathcal{N}(]0,x])})$ denotes the expected value of the Poisson transform of the sequence (R_n), then Equation (6) gives the relation

$$R(x) = 1 + \sum_{\ell=2}^{+\infty} \mathbb{P}(G = \ell)\, \mathbb{E}\left(\sum_{i=1}^{\ell} R(xV_{i,\ell})\right) - \mathbb{E}(G)\mathbb{P}(t_D \geq x),$$

by denoting

$$h(x) = 1 - \mathbb{E}(G) \int_x^{+\infty} \frac{u^{D-1}}{(D-1)!}\, du,$$

it is easy to see that the above identity can be written as the following integral equation

$$R(x) = \int_0^{+\infty} R(xu)\, \frac{\mathcal{W}(du)}{u} + h(x). \tag{14}$$

Recall that $\mathcal{W}$ is some probability distribution on the interval $[0,1]$. For the Q-ary protocol considered by Mathys and Flajolet [29], this equation is

$$R(x) = \sum_{i=1}^{Q} R(xp_i) + h(x).$$

It is analyzed by considering the Mellin transform $R^*(s)$ of $R(x)$ on some vertical strip $\mathcal{S}$ of $\mathbb{C}$,

$$R^*(s) = \int_0^{+\infty} R(u) u^{s-1}\, du, \quad s \in \mathcal{S}$$

which, in this case, is given by

$$R^*(s) = h^*(s) \Bigg/ \left(1 - \sum_{i=1}^{Q} \frac{1}{p_i^s}\right).$$

The analytical approach consists in analyzing the poles of $R^*(s)$ on the right hand side of $\mathcal{S}$, basically the solutions with positive real part of the equation

$$p_1^{-s} + p_2^{-s} + \cdots + p_Q^{-s} = 1.$$

Then, by inverting the Mellin transform and using complex analysis techniques, the asymptotic behavior of $(R(x))$ at infinity is described in terms of these poles. The final step, an analytic inversion of the Poisson transform together with technical estimates, establishes a relation between the asymptotic behaviors of the function $x \to R(x)$ and of the sequence (R_n).

In the general case considered here, Equation (14) gives the following expression for the Mellin transform of $(R(x))$

$$R^*(s) = h^*(s) \Bigg/ \left(1 - \int_0^{+\infty} \frac{1}{u^{s+1}}\, \mathcal{W}(du)\right)$$

An analogue of the analytic approach would start with the study of the roots $s \in \mathbb{C}$, $\Re(s) \geq 0$, of the equation

$$\int_0^{+\infty} \frac{1}{u^{s+1}}\, \mathcal{W}(du) = 1, \tag{15}$$

and, if possible, proceeds with successive inversions of Mellin transform and Poisson transform.

As it will be seen, our direct approach reduces to the minimum the technical apparatus required for such an analysis. The Poisson transform of (R_n) is also used in our method, but it is conveniently represented, see Equation (8), so that it can be right away inverted to give an explicit expression (12) for $\mathbb{E}(R_n)$ which will give directly the asymptotic behavior of the sequence $(\mathbb{E}(R_n))$.

Interestingly, if (L_n) denotes the non-increasing sequence of the lengths of the subintervals of $[0, 1]$ associated to the splitting procedure (see Section 1.3). The *Zeta function* of the string (L_n) is defined as the meromorphic function

$$\zeta(s) = \sum_{n \geq 1} L_n^s, \quad s \in \mathbb{C},$$

see Hambly and Lapidus [14] and Lapidus and van Frankenhuysen [23]. It is not difficult to see that the relation

$$\mathbb{E}(\zeta(s)) = \int_0^{+\infty} u^s\, \mathcal{W}(du) \Bigg/ \left(1 - \int_0^{+\infty} u^s\, \mathcal{W}(du)\right)$$

holds. In particular, the poles of the Zeta function of the associated random recursive string can be expressed in terms of the solutions of Equation (15).

3. Analysis of the Asymptotic Average Cost

An Associated Random Walk. If (W_i) is an i.i.d. sequence with common distribution $\mathcal{W}$ defined by Equation (1), the sequence $(B_i) = (-\log(W_i))$ is an i.i.d. sequence of non-negative random variables. The random walk (S_n) associated to (B_i),

$$S_n = B_1 + B_2 + \cdots + B_n, \qquad n \geq 0.$$

As it will be seen, the asymptotic behavior of the splitting algorithm depends much on the distribution of (B_i). For $x > 0$, the crossing time ν_x of level x by (S_n) is defined as

$$\nu_x = \inf\{n : S_n > x\}.$$

For $0 < y < 1$, the variable $T(y)$ of Proposition 6 is simply $\nu_{-\log(y)}$. Section 5 recalls the main results concerning Renewal Theory used in the following.

If Ψ is defined as

$$\Psi(x) = \mathbb{E}\left(\sum_{i=0}^{\nu_x - 1} \exp\left(\sum_{k=1}^{i} B_k\right)\right), \quad x > 0$$

then by Equation (12),

$$\mathbb{E}(R_n) = 1 + \mathbb{E}(G)\mathbb{E}\left[\Psi\left(-\log\left(U^D_{(n)}\right)\right)\right]. \tag{16}$$

It is clear that $-\log(U^D_{(n)})$ converges in distribution to $+\infty$ as n goes to infinity. The asymptotic behavior of Ψ at infinity is first analyzed, this function can be rewritten as

$$\Psi(x)e^{-x} = \mathbb{E}\left(\sum_{i=0}^{\nu_x-1} e^{S_i-x}\right) = \mathbb{E}\left(\sum_{i=1}^{\nu_x} e^{S_{\nu_x-i}-x}\right). \tag{17}$$

3.1. **The Non-Arithmetical Case.** In this part, it is assumed that the distribution of W_1 is not exponentially arithmetic. See Definition 2.

Lemma 8. *Under the condition*

$$\mathbb{E}\left(\frac{|\log(W_1)|}{W_1}\right) = \int_0^1 \frac{|\log(x)|}{x}\,\mathcal{W}(dx) < +\infty,$$

the relation

$$\sup_{x\geq 0} \mathbb{E}\left(e^{S_{\nu_x}-x}\right) < +\infty$$

holds.

Proof. Lorden's Inequalities, see Lorden [25] and Chang [5], show that, for any $p \geq 0$,

$$\sup_{x\geq 0} \mathbb{E}\left((S_{\nu_x}-x)^p\right) \leq \frac{p+2}{(p+1)\mathbb{E}(B_1)}\mathbb{E}\left(B_1^{p+1}\right),$$

thus, one gets the relation

$$\begin{aligned}\sup_{x\geq 0} \mathbb{E}\left(e^{S_{\nu_x}-x}\right) &\leq \frac{1}{\mathbb{E}(B_1)}\int_0^{+\infty} (u+2)e^u\,\mathbb{P}(B_1\geq u)\,du \\ &= \mathbb{E}\left((B_1+1)e^{B_1}\right) - 1 = \mathbb{E}\left(\frac{-\log(W_1)+1}{W_1}\right) - 1 < +\infty.\end{aligned}$$

□

For $i > 1$, Theorem 20 shows that, when x goes to infinity, the variable $S_{\nu_x-i}-x$ converges in distribution to $-(\tau^*+\tau_1+\tau_2+\cdots+\tau_{i-1})$, where the variables (τ_n) are i.i.d. distributed as B_1 and independent of τ^* whose distribution is given by

$$\mathbb{E}(f(\tau^*)) = \frac{1}{\mathbb{E}(B_1)}\int_0^{+\infty} f(u)\mathbb{P}(B_1\geq u)\,du,$$

for any non-negative Borelian function on $\mathbb{R}$. By Assumption (A), the increments of the random walk (S_n) are bounded below by $-\log(\delta)$, therefore one gets the relation, for $1 < K \leq \nu_x$,

$$\sum_{i=K}^{\nu_x} e^{S_{\nu_x-i}-x} = e^{S_{\nu_x}-x}\sum_{i=K}^{\nu_x} e^{S_{\nu_x-i}-S_{\nu_x}} \leq e^{S_{\nu_x}-x}\,\frac{\delta^K}{1-\delta}.$$

From Lemma 8 and Equation (17), one deduces then

$$\lim_{x\to+\infty}\Psi(x)e^{-x}=\mathbb{E}\left(\sum_{i=1}^{+\infty}\exp\left(-\tau^*-\tau_1-\tau_2-\cdots-\tau_{i-1}\right)\right)$$

$$=\frac{1-\mathbb{E}\left(\exp(-\tau_1)\right)}{\mathbb{E}(\tau_1)}\times\frac{1}{1-\mathbb{E}\left(\exp(-\tau_1)\right)}=\frac{1}{-\mathbb{E}(\log(W_1))},\tag{18}$$

since, by Equation (31), the density of τ^* on $\mathbb{R}_+$ is given by

$$\mathbb{P}(\tau_1\geq x)/\mathbb{E}(\tau_1),\quad x\geq 0.$$

Proposition 9 (Convergence of Averages)**.** *If the distribution of W_1 is not exponentially arithmetic and such that*

$$\mathbb{E}\left(\frac{|\log(W_1)|}{W_1}\right)<+\infty,$$

then the following convergence holds

$$\lim_{n\to+\infty}\frac{\mathbb{E}(R_n)}{n}=\frac{\mathbb{E}(G)}{(D-1)\mathbb{E}(-\log W_1)}.$$

Proof. Equation (16) gives that, for $n\geq 1$,

$$\frac{\mathbb{E}(R_n)}{n}=\frac{1}{n}+\mathbb{E}(G)\mathbb{E}\left(\Psi\left[-\log\left(U_{(n)}^D\right)\right]\exp\left(\log\left(U_{(n)}^D\right)\right)\frac{1}{nU_{(n)}^D}\right).$$

As n goes to infinity the variable $nU_{(n)}^D$ converges in distribution to a random variable t_D which is a sum of D i.i.d. exponential random variables with parameter 1, furthermore,

$$\lim_{n\to+\infty}\mathbb{E}\left(\frac{1}{nU_{(n)}^D}\right)=\mathbb{E}(1/t_D)=\frac{1}{D-1}.$$

For $\varepsilon>0$, there exists K such that, for $x>K$, $|\Psi(x)\exp(-x)+1/\mathbb{E}(\log W_1)|<\varepsilon$, if C denotes the supremum of $x\to\Psi(x)\exp(-x)$ on $\mathbb{R}_+$, then

$$\left|\mathbb{E}\left(\Psi\left[-\log\left(U_{(n)}^D\right)\right]\exp\left(\log\left(U_{(n)}^D\right)\right)\frac{1}{nU_{(n)}^D}\right)-\frac{1}{(D-1)\mathbb{E}(-\log W_1)}\right|\tag{19}$$
$$\leq\varepsilon\mathbb{E}\left(\frac{1}{nU_{(n)}^D}\right)+\left(C+\frac{1}{\mathbb{E}(-\log W_1)}\right)\mathbb{E}\left(\mathbb{1}_{\{U_{(n)}^D>\exp(-K)\}}\frac{1}{nU_{(n)}^D}\right)$$
$$+\frac{1}{\mathbb{E}(-\log W_1)}\left|\mathbb{E}\left(\frac{1}{nU_{(n)}^D}\right)-\frac{1}{D-1}\right|.$$

For $K_2>0$,

$$\limsup_{n\to+\infty}\mathbb{E}\left(\mathbb{1}_{\{U_{(n)}^D>\exp(-K)\}}\frac{1}{nU_{(n)}^D}\right)$$
$$\leq\limsup_{n\to+\infty}\mathbb{E}\left(\mathbb{1}_{\{nU_{(n)}^D>K_2\exp(-K)\}}\frac{1}{nU_{(n)}^D}\right)=\mathbb{E}\left(\mathbb{1}_{\{t_D>K_2\exp(-K)\}}\frac{1}{t_D}\right),$$

and this term goes to 0 as K_2 tends to infinity. One concludes that the right hand side of Relation (19) is arbitrarily small as n goes to infinity. The proposition is proved. □

Corollary 10.

(1) *Q-ary protocol with blocked arrivals.*
When $D=2$, $G\equiv Q$ and $V_{i,Q}=p_i$ for $1\leq i\leq Q$ then, if at least one of the real numbers $\log p_i/\log p_1$, $2\leq i\leq Q$ is not rational, the convergence
$$\lim_{n\to+\infty}\frac{\mathbb{E}(R_n)}{n}=\frac{Q}{\sum_{i=1}^{Q}-p_i\log p_i}.$$
holds.

(2) *Symmetrical case.*
If G is not a degenerated random variable such that $\mathbb{E}(\log G)<+\infty$ and, for $\ell\geq 2$ and $1\leq i\leq \ell$, $V_{i,\ell}=1/\ell$, then
$$\lim_{n\to+\infty}\frac{\mathbb{E}(R_n)}{n}=\frac{\mathbb{E}(G)}{(D-1)\mathbb{E}(\log G)}.$$

3.2. **The Arithmetical Case.** It is assumed that the distribution of W_1 is exponentially arithmetic with exponential span $\lambda>0$. The law of $-\log(W_1)/\lambda$ is a probability distribution on $\mathbb{N}$. For $i\geq 1$, one defines $C_i=B_i/\lambda=-\log(W_i)/\lambda$. In the arithmetic case, the integer valued random walk associated to (C_i) plays the key role, much in the same way as for (S_n) in the non-arithmetic case. By denoting
$$\tau_n=\inf\left\{k\geq 1:\sum_{i=1}^{k}C_i\geq n\right\},$$
Equation (17) can be rewritten as, for $x\geq 0$,
$$\Psi(x)e^{-\lambda\lceil x/\lambda\rceil}=\mathbb{E}\left[\sum_{i=1}^{\tau_{\lceil x/\lambda\rceil}}\exp\left(\lambda\left(\sum_{k=1}^{\tau_{\lceil x/\lambda\rceil}-i}C_k-\lceil x/\lambda\rceil\right)\right)\right],$$
where $\lceil y\rceil=\inf\{n\in\mathbb{N}:n>y\}$ for $y\geq 0$. By using Theorem 21 and its notations, for $i\geq 1$, as n goes to infinity, the variable $C_1+\cdots+C_{\tau_n-i}-n$ converges in distribution to $-(C_1^*+C_2+\cdots+C_i)$. With the same method as in the non-arithmetic case, if the variable $|\log(W_1)|/W_1$ is integrable, then
$$\lim_{x\to+\infty}\Psi(x)e^{-\lambda\lceil x/\lambda\rceil}=\frac{1}{\mathbb{E}(|\log(W_1)|)}\,\frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}.$$

Proposition 11 (Asymptotic Periodic Oscillations)**.** *If the distribution of W_1 is exponentially arithmetic with exponential span $\lambda>0$, and such that*
$$\mathbb{E}\left(\frac{|\log(W_1)|}{W_1}\right)<+\infty,$$
then, as n gets large, the equivalence
$$\frac{\mathbb{E}(R_n)}{n}\sim F\left(\frac{\log n}{\lambda}\right)$$
holds, where F is the periodic function with period 1 defined by, for $x\geq 0$,
$$F(x)=\frac{\mathbb{E}(G)}{\mathbb{E}(|\log(W_1)|)}\,\frac{\lambda}{1-e^{-\lambda}}\int_0^{+\infty}\exp\left(-\lambda\left\{x-\frac{\log y}{\lambda}\right\}\right)\frac{y^{D-2}}{(D-1)!}e^{-y}\,dy$$
and $\{x\}=x-\lfloor x\rfloor$.

Proof. For $n \geq 1$, if $\lceil x \rceil = \lfloor x \rfloor + 1$,

$$\frac{1}{n}\mathbb{E}\left[\Psi\left(-\log(U^D_{(n)})\right)\right]$$
$$= \mathbb{E}\left[\Psi\left(-\log U^D_{(n)}\right) e^{-\lambda\lceil -\log(U^D_{(n)}/\lambda)\rceil} e^{\lambda} \exp\left(-\lambda\left\{\frac{-\log(U^D_{(n)})}{\lambda}\right\}\right)\frac{1}{nU^D_{(n)}}\right],$$

since $nU^D_{(n)}$ converges in distribution to t_D as n goes to infinity, with the same method as in the proof of Proposition 9, one gets the equivalences

$$\frac{1}{n}\mathbb{E}\left[\Psi\left(-\log(U^D_{(n)})\right)\right] \times \mathbb{E}(|\log(W_1)|)\,\frac{1-e^{-\lambda}}{\lambda}$$
$$\sim E\left[\exp\left(-\lambda\left\{\frac{\log(n)}{\lambda} - \frac{\log(nU^D_{(n)})}{\lambda}\right\}\right)\frac{1}{nU^D_{(n)}}\right]$$
$$= E\left[\exp\left(-\lambda\left\{\frac{\log(n)}{\lambda} - \frac{\log t_D}{\lambda}\right\}\right)\frac{1}{t_D}\right].$$

One concludes by using Equation (16). □

Corollary 12 (Q-ary protocol with blocked arrivals)**.** *When $D = 2$, $G \equiv Q$ and $V_{i,Q} = p_i$ for $1 \leq i \leq Q$ then, if all the real numbers $\log p_i/\log p_1$, $2 \leq i \leq Q$ are rational, the equivalence*

$$\frac{\mathbb{E}(R_n)}{n} \sim F\left(\frac{\log n}{\lambda}\right)$$

holds, where F is the periodic function with period 1 defined by, for $x \geq 0$,

$$F(x) = \frac{Q}{-\sum_{i=1}^{Q} p_i \log p_i}\,\frac{\lambda}{1-e^{-\lambda}}\int_0^{+\infty}\exp\left(-\lambda\left\{x - \frac{\log y}{\lambda}\right\}\right)e^{-y}\,dy,$$

where $\{x\} = x - \lfloor x \rfloor$ and $\lambda = \sup\{y > 0 : \forall i \in \{1,\ldots,Q\}, \log p_i \in y\mathbb{Z}\}$.

4. The Distributions of the Symmetrical Q-Ary Algorithm

From now on, it is assumed that the branching degree of the splitting algorithm is constant, i.e. $\mathbb{P}(G = Q) = 1$, and uniform, $V_{i,Q} \equiv 1/Q$ for $1 \leq i \leq Q$. A group of $n \geq D$ items is randomly, equally divided into Q subgroups. From Proposition 11, it is known that

$$\mathbb{E}(R_n)/n \sim F_1\left(\log_Q n\right)$$

as n goes to infinity, with

$$F_1(x) = \frac{Q^2}{Q-1}\int_0^{+\infty} Q^{-\left\{x-\log_Q y\right\}}\frac{y^{D-2}}{(D-1)!}\,e^{-y}\,dy. \tag{20}$$

This is a, typical, case where a regular law of large numbers does not hold.

The purpose of this section is to strengthen the above convergence. The distribution of the Poisson transform of the sequence (R_n), i.e. the random variable $R_{\mathcal{N}(]0,x])}$, is investigated and not only its average as before. In particular it is shown that, for the Poisson transform, a *standard* law of large numbers can be used to prove the oscillating behavior of the algorithm. In other words, these uncommon laws of large numbers can be, in the end, expressed in a classical probabilistic setting.

Notations. Throughout the rest of the paper it is assumed that

(1) $\mathcal{N}$ is a Poisson process with intensity 1 on $\mathbb{R}_+$. Another Poisson process will be used but in the two dimensional space $[0,1]\times\mathbb{R}_+$.
(2) The variable $\mathcal{M}$ denotes a Poisson process on $[0,1]\times\mathbb{R}_+$ with intensity 1, this is a distribution of random points on $[0,1]\times\mathbb{R}_+$ with the following properties: if $\mathcal{M}(H)$ denotes the number of points that "fall" into the set $H\subset[0,1]\times\mathbb{R}_+$,
   — For $x\in[0,1]\times\mathbb{R}_+$, $M(\{x\})\in\{0,1\}$
   — If G and H are disjoint subsets of $[0,1]\times\mathbb{R}_+$, the variables $\mathcal{M}(G)$ and $\mathcal{M}(H)$ are independent.
   — The distribution of the variable $\mathcal{M}([a,b]\times[y,z])$ is Poisson with parameter $(b-a)(z-y)$ for $0\leq a\leq b\leq 1$ and $0\leq y\leq z$.

   Note that the random variables $\mathcal{N}([0,x])$ and $\mathcal{M}([0,1]\times[0,x])$ have a Poisson distribution with parameter x.
(3) The Poisson transform of the sequence (R_n) is denoted by $\mathcal{R}(x)$, $x\geq 0$,

$$\mathcal{R}(x)\stackrel{\text{dist.}}{=}R_{\mathcal{N}([0,x])}\stackrel{\text{dist.}}{=}R_{\mathcal{M}([0,1]\times[0,x])}.$$

Its expectation is given by Equation (8). This section is devoted to the study of the asymptotic behavior of the *distribution* of $\mathcal{R}(x)$.

4.1. **Laws of Large Numbers.** In this section, it is proved that the Poisson transform of the sequence (R_n) satisfies a strong law of large numbers. A nice representation of this transform as a functional of Poisson processes is first proved in the following proposition.

Proposition 13. *The distribution of the Poisson transform $\mathcal{R}(x)$ of the sequence (R_n) satisfies the following relations,*

$$\mathcal{R}(x)\stackrel{dist.}{=}\mathcal{R}_1(x)\stackrel{def.}{=}1+Q\sum_{p\geq 0}\sum_{k=0}^{Q^p-1}\phi_{\mathcal{N}}\left(xk/Q^p,x(k+1)/Q^p\right),\tag{21}$$

where, for $0\leq a\leq b$ $\phi_{\mathcal{N}}(a,b)=1$ if $\mathcal{N}(]a,b])\geq D$ and 0 otherwise,

$$\mathcal{R}(x)\stackrel{dist.}{=}\mathcal{R}_2(x)\stackrel{def.}{=}1+Q\sum_{p\geq 0}\sum_{k=0}^{Q^p-1}\mathbb{1}_{\{\mathcal{M}(]k/Q^p,(k+1)/Q^p]\times[0,x])\geq D\}}.\tag{22}$$

Note that the function $x\to\mathcal{R}_2(x)$ is clearly non-decreasing. In particular, if f is some non-decreasing function on $\mathbb{R}_+$, the same property holds for

$$x\to\mathbb{E}[f(\mathcal{R}(x))].$$

Representation (21) will be useful to get a strong law of large numbers on subsequences and Representation (21) will be used to get the full convergence in distribution of $\mathcal{R}(x)/x$ as x tends to infinity.

Proof. By the splitting property of Poisson random variables, the recurrence relation (2) for the sequence (R_n) can be expressed as

$$R_{\mathcal{N}(]0,x])}\stackrel{\text{dist.}}{=}1+\sum_{i=1}^{Q}R_{i,\mathcal{N}(]x(i-1)/Q,xi/Q])}-Q\mathbb{1}_{\{\mathcal{N}(]0,x])<D\}},$$

for $x \geq 0$. If, for $0 \leq a < b$,

$$\Phi(a,b) = \frac{1}{Q}(R_{\mathcal{N}(]a,b])} - 1), \tag{23}$$

the last equation can be rewritten as

$$\Phi(0,x) = \sum_{i=1}^{Q} \Phi_i\left(\frac{i-1}{Q}\,x, \frac{i}{Q}\,x\right) + \phi_{\mathcal{N}}(0,x), \tag{24}$$

with an obvious notation with the subscripts i for Φ. By iterating this relation, one gets that, almost surely, the expansion

$$\Phi(0,x) = \sum_{p\geq 0}\sum_{k=0}^{Q^p-1} \phi_{\mathcal{N}}\left(\frac{k}{Q^p}x, \frac{k+1}{Q^p}x\right), \tag{25}$$

holds. The function $\Phi(0,x)$ is just the sum of the function $\phi_{\mathcal{N}}$ on the Q-adic intervals of $[0,x]$.

Equation (22) is proved in the same way. □

Representation of some of the functionals of the associated tree. When $\mathcal{N}([0,x])$ items are at the root of the associated tree, the total number of node of the tree $R_{\mathcal{N}([0,x])}$ is not the only quantity that can be represented, by Equation (21) in terms of the Poisson process $\mathcal{N}$.

The *maximal depth* $M(x)$ of the associated tree when there are $\mathcal{N}([0,x])$ items at the top of the tree can be expressed as a functional of the Poisson process

$$M(x) = \max\{p \geq 1 : \exists k, 0 \leq k < Q^{p-1}-1, \mathcal{N}(]k/Q^{p-1},(k+1)/Q^{p-1}]) \geq D\}$$

The quantity

$$F(x) = \max\{p \geq 1 : \forall k, 0 \leq k < Q^{p-1}-1, \mathcal{N}(]k/Q^{p-1},(k+1)/Q^{p-1}]) \geq D\}$$

is the *number of full levels* of the tree. See Knessl and Szpankowski [21]. Note that these quantities are directly related to classical occupancy problems. The *number of nodes at level* $p \geq 1$ is given by

$$Q\sum_{k=0}^{Q^{p-1}-1} \mathbb{1}_{\{\mathcal{N}(]k/Q^{p-1},(k+1)/Q^{p-1}]) \geq D\}}.$$

This is not, of course, an exhaustive list of the possible representations in terms of the Poisson process.

It is quite useful to think splitting algorithms either in terms of trees or in terms of Q-adic subintervals of $[0,1]$. In a more general case, i.e. when the splitting algorithm is not symmetrical, a representation similar to Representation (21) can be obtained by using the associated random decomposition of the interval $[0,1]$ instead of the Q-adic decomposition. See Falconer [10].

A strong law of large numbers. Equation (24) shows that, if $N > 0$, the quantity $\Phi\left(0, yQ^N\right)$ is the sum of the Φ on the intervals $[yp, yp+y]$, $0 \leq p < Q^N$, and of $\phi_{\mathcal{N}}$ on the intervals $[ykQ^n, y(k+1)Q^n]$ contained in $[0, yQ^N]$, that is

$$\Phi\left(0, yQ^N\right) = \sum_{p=0}^{Q^N-1} \Phi(yp, yp+y) + \sum_{n=1}^{N}\sum_{k=0}^{Q^{N-n}-1} \phi_{\mathcal{N}}\left(ykQ^n, y(k+1)Q^n\right). \tag{26}$$

By the independence properties of Poisson process, the classical strong law of large numbers shows that, almost surely,

$$\lim_{N\to+\infty}\frac{1}{Q^N}\sum_{p=0}^{Q^N-1}\Phi(yp,yp+y)=\mathbb{E}(\Phi(0,y))=\sum_{p\geq 0}Q^p\,\mathbb{E}(\phi_{\mathcal{N}}\left(0,y/Q^p\right)$$
$$=\sum_{p\geq 0}Q^p\mathbb{P}\left(\mathcal{N}(]0,y/Q^p])\geq D\right),$$

by using Equation (25), and for $n>0$,

$$\lim_{N\to+\infty}\frac{1}{Q^N}\sum_{k=0}^{Q^{N-n}-1}\phi_{\mathcal{N}}\left(ykQ^n,y(k+1)Q^n\right)=\frac{1}{Q^n}\mathbb{E}(\phi_{\mathcal{N}}(0,yQ^n)).$$

Note that, for $0<K<N$,

$$\sum_{n=K}^{N}\frac{1}{Q^N}\sum_{k=0}^{Q^{N-n}-1}\phi_{\mathcal{N}}\left(ykQ^n,y(k+1)Q^n\right)\leq\sum_{n=K}^{N}\frac{1}{Q^N}Q^{N-n}\leq\frac{1}{Q^{K-1}}.$$

The three last identities and Decomposition (26) give that, almost surely,

$$\lim_{N\to+\infty}\frac{1}{yQ^N}\Phi\left(0,yQ^N\right)=\sum_{n\in\mathbb{Z}}\frac{1}{yQ^n}\mathbb{P}(\mathcal{N}(]0,yQ^n])\geq D).$$

Proposition 14 (Strong Law of Large Numbers)**.** *With the same notations as in Proposition 13, for $0<y\leq Q$, almost surely,*

$$\lim_{N\to+\infty}\frac{\mathcal{R}_1\left(yQ^N\right)}{yQ^N}=Q\sum_{n\in\mathbb{Z}}\frac{1}{yQ^n}\mathbb{P}(\mathcal{N}(]0,yQ^n])\geq D)\tag{27}$$
$$=Q\sum_{n\in\mathbb{Z}}\frac{1}{yQ^n}\int_0^{yQ^n}\frac{u^{D-1}}{(D-1)!}e^{-u}\,du=F_1(\log_Q y),$$

where F_1 is the periodic function defined by Equation (20)*.*

As a byproduct, the proposition establishes the intuitive (and classical) fact that the sequence $(\mathbb{E}(R_n)/n)$ and the function $x\to\mathbb{E}(R_{\mathcal{N}([0,x])})/x$ have the same asymptotic behavior at infinity. Note that if $G(y)$ is defined as the second term of Equation (27), then the function $x\to G(Q^x)$ is clearly periodic with period 1.

Proof. Clearly, only the relation $F_1(\log_Q y)=G(y)$ has to be proved. For $n\in\mathbb{Z}$, if t_D is the Dth point of the Poisson process $\mathcal{N}$ then

$$\mathbb{P}(\mathcal{N}(]0,yQ^n])\geq D)=\mathbb{P}(t_D\leq yQ^n)=\int_0^{+\infty}\mathbb{1}_{\{u\leq yQ^n\}}\frac{u^{D-1}}{(D-1)!}\,e^{-u}\,du.$$

By summing up these terms, with Fubini's Theorem one gets

$$G(y)=Q\int_0^{+\infty}\sum_{n\in\mathbb{Z}}\frac{1}{yQ^n}\mathbb{1}_{\{u\leq yQ^n\}}\frac{u^{D-1}}{(D-1)!}\,e^{-u}\,du$$
$$=\frac{Q^2}{Q-1}\int_0^{+\infty}\frac{1}{yQ^{\lceil\log_Q(u/y)\rceil}}\frac{u^{D-1}}{(D-1)!}\,e^{-u}\,du$$
$$=\frac{Q^2}{Q-1}\int_0^{+\infty}\frac{1}{Q^{-\{\log_Q(u/y)\}}}\frac{u^{D-2}}{(D-1)!}\,e^{-u}\,du=F_1(\log_Q y).$$

The proposition is proved. □

The following proposition establishes a weak law of large number for the Poisson transform of the sequence (R_n). Devroye [8] obtained related results in a more general framework by using Talagrand's concentration inequalities.

Theorem 15 (Law of Large Numbers)**.** *The following convergence in distribution holds, for any* $\varepsilon > 0$*,*

$$\lim_{x\to+\infty} \mathbb{P}\left(\left|\frac{\mathcal{R}(x)}{xF_1(\log_Q x)} - 1\right| \geq \varepsilon\right) = 0,$$

where F_1 *is the function defined by Equation* (20)*.*

Proof. For $x > 0$, one defines $N_x = \lfloor \log_Q x\rfloor$, $u_x = x/Q^{N_x}$ and, for $p \geq 1$, $z_x = \lfloor u_x p\rfloor Q^{N_x}/p$. Note that $\sup_{x\geq 1} |x/z_x - 1|$ converges to 0 as p tends to infinity, hence by continuity of F_1,

$$\lim_{p\to+\infty} \sup_{x\geq 1} \left|\frac{xF_1(\log_Q x)}{z_xF_1(\log_Q z_x)} - 1\right| = 0.$$

Proposition 14 shows that for $p \geq 1$, almost surely, for k, $0 \leq k \leq p$,

$$\lim_{N\to+\infty} \frac{\mathcal{R}_1(y_kQ^N)}{y_kQ^NF_1(\log_Q y_k)} = 1,$$

with $y_k = k/p$. Therefore, if $p \geq 1$ is fixed, almost surely,

$$\lim_{x\to+\infty} \frac{\mathcal{R}_1(z_x)}{z_xF_1(\log_Q z_x)} = 1,$$

The monotonicity of the function $x \to \mathcal{R}_2(x)$ gives the relation

$$\mathcal{R}_2\left(\frac{\lfloor u_xp\rfloor}{p}Q^{N_x}\right) \leq \mathcal{R}_2(x),$$

one finally gets

$$\begin{aligned}\mathbb{P}&\left(\frac{\mathcal{R}(x)}{xF_1(\log_Q x)} < 1-\varepsilon\right) = \mathbb{P}\left(\frac{\mathcal{R}_2(x)}{xF_1(\log_Q x)} < 1-\varepsilon\right)\\ &\leq \mathbb{P}\left(\frac{\mathcal{R}_2(z_x)}{xF_1(\log_Q x)} < 1-\varepsilon\right) = \mathbb{P}\left(\frac{\mathcal{R}_1(z_x)}{z_xF_1(\log_Q z_x)}\frac{z_xF_1(\log_Q z_x)}{xF_1(\log_Q x)} < 1-\varepsilon\right),\end{aligned}$$

therefore,

$$\lim_{x\to+\infty} \mathbb{P}\left(\frac{\mathcal{R}(x)}{xF_1(\log_Q x)} < 1-\varepsilon\right) = 0.$$

The analogous inequality is obtained in the same way. The theorem is proved. □

4.2. **Central Limit Theorems.** For $N \geq 1$ and $0 < x \leq Q$, with Φ defined by Equation (23), the variance of the variable $\Phi(0, x)$ is first analyzed. The expansion (25) gives

$$[\Phi(0,x) - \mathbb{E}(\Phi(0,x))]^2 = \sum_{p\geq 0}\sum_{k=0}^{Q^p-1}\sum_{p'\geq 0}\sum_{k'=0}^{Q^{p'}-1} \Delta_{k,p}(x)\Delta_{k',p'}(x),$$

with

$$\Delta_{k,p}(x) = \phi_{\mathcal{N}}\left(\frac{k}{Q^p}x, \frac{k+1}{Q^p}x\right) - \mathbb{E}\left(\phi_{\mathcal{N}}\left(0, \frac{1}{Q^p}x\right)\right).$$

The expected value of the variable $\Delta_{k,p}(x)\Delta_{k',p'}(x)$ is non-zero only if $p \leq p'$ and $kQ^{p'-p} \leq k' \leq (k+1)Q^{p'-p}-1$ or the symmetrical condition by exchanging (p,k) and (p',k').

$$\mathbb{E}\left[\Big(\Phi(0,x)-\mathbb{E}\left(\Phi(0,x)\right)\Big)^2\right] = \sum_{p\geq 0}\sum_{k=0}^{Q^p-1}\mathbb{E}\left[\Delta_{k,p}(x)^2\right] + 2\sum_{p\geq 0}\sum_{k=0}^{Q^p-1}\sum_{p'>p}\sum_{k'=kQ^{p'-p}}^{(k+1)Q^{p'-p}-1}\mathbb{E}\left[\Delta_{k,p}(x)\Delta_{k',p'}(x)\right].$$

By using the elementary identities

$$\begin{aligned}
\mathbb{E}\left[\Delta_{k,p}(x)^2\right] &= \mathbb{E}\left[\phi_{\mathcal{N}}(0,x/Q^p)\right]\left(1-\mathbb{E}\left[\phi_{\mathcal{N}}(0,x/Q^p)\right]\right)\\
&= \mathbb{P}(t_D \leq x/Q^p)\mathbb{P}(s_D \geq x/Q^p),\\
\mathbb{E}\left[\Delta_{k,p}(x)\Delta_{k',p'}(x)\right] &= \mathbb{E}\left[\phi_{\mathcal{N}}(0,x/Q^{p'})\right]\left(1-\mathbb{E}\left[\phi_{\mathcal{N}}(0,x/Q^p)\right]\right)\\
&= \mathbb{P}(t_D \leq x/Q^{p'})\mathbb{P}(s_D \geq x/Q^p),
\end{aligned}$$

where t_D and s_D are independent random variables with the same distribution as the Dth point of the Poisson process $\mathcal{N}$, one gets the relation

$$\mathbb{E}\left[\Big(\Phi(0,x)-\mathbb{E}\left(\Phi(0,x)\right)\Big)^2\right] = \sum_{p\geq 0}Q^p\mathbb{P}(t_D \leq x/Q^p \leq s_D) + 2\sum_{p'>p\geq 0}Q^{p'}\mathbb{P}(t_D \leq x/Q^{p'}, x/Q^p \leq s_D).$$

By switching again the series and the expected values, one finally obtains,

$$\begin{aligned}
(28)\quad &(Q-1)\mathbb{E}\left[\Big(\Phi(0,x)-\mathbb{E}\left(\Phi(0,x)\right)\Big)^2\right]\\
&= Q\mathbb{E}\left(\left(Q^{\lfloor \log_Q(x/t_D)\rfloor}-Q^{\lfloor \log_Q(x/s_D)\rfloor}\right)\mathbb{1}_{\{\lfloor \log_Q(x/t_D)\rfloor > \lfloor \log_Q(x/s_D)\rfloor\}}\right)\\
&\quad + 2Q\mathbb{E}\left(\left(\lfloor \log_Q(x/t_D)\rfloor - \lfloor \log_Q(x/s_D)\rfloor -1\right)^+ Q^{\lfloor \log_Q(x/t_D)\rfloor}\right)\\
&\quad - 2\frac{Q}{Q-1}\mathbb{E}\left(\left(Q^{\lfloor \log_Q(x/t_D)\rfloor}-Q^{\lfloor \log_Q(x/s_D)\rfloor+1}\right)^+\right),
\end{aligned}$$

where $a^+ = \max(a,0)$ for $a \in \mathbb{R}$. This identity gives the following proposition. A similar proposition has been proved by Jacquet and Régnier [16] and Régnier and Jacquet [37] in the case where $Q = D = 2$ but without symmetry conditions as it is the case here. See also Mahmoud [27] Chapter 5.

Proposition 16 (Asymptotic Variance)**.** *The variance of the Poisson transform of the sequence (R_n) satisfies the following equivalence, as x goes to infinity,*

$$\frac{1}{x}\operatorname{Var}\left(\mathcal{R}(x)\right) \sim F_2(\log_Q(x)),$$

where F_2 is the continuous periodic function with period 1 defined by, for $y \geq 0$,

$$F_2(y) = \int_{\mathbb{R}_+^2} f_2\left(\{y - \log_Q(u)\}, \{y - \log_Q(v)\}, u, v\right) \times \frac{u^{D-1}}{(D-1)!}\frac{v^{D-1}}{(D-1)!}e^{-(u+v)}\, dudv, \tag{29}$$

with $\{z\} = z - \lfloor z \rfloor$ for $z \in \mathbb{R}$ and for $u > 0$, $v > 0$ and $y \in \mathbb{R}$,

$$f_2(a,b,u,v) = \frac{Q}{Q-1}\left(\frac{Q^a}{u} - \frac{Q^b}{v}\right)\mathbb{1}_{\{\log_Q(v/u)+b>a\}} + \frac{2Q}{Q-1}\left(\log_Q(v/u) - a + b - 1\right)^+ - \frac{2Q}{(Q-1)^2}\left(\frac{Q^a}{u} - \frac{Q^{b+1}}{v}\right)^+$$

where $z^+ = \max(z, 0)$.

Note that a more detailed expansion of the variance could be obtained with Formula (28).

Proposition 17 (Central Limit Theorem for Poisson Transform)**.** *For $0 < y < Q$, as N tends to infinity, the variable*

$$\frac{1}{\sqrt{Q^N}}\left(\mathcal{R}(yQ^N) - \mathbb{E}(\mathcal{R}(yQ^N))\right)$$

converges in distribution to a Gaussian centered random variable with variance $yF_2(\log_Q y)$, where F_2 is defined by Equation (29).

Proof. It is enough to prove the proposition for the variable $\Phi(0,x)$ defined by

$$\Phi(0,x) = \frac{1}{Q}(R_{\mathcal{N}(]0,x])} - 1),$$

Equation (26) gives, for $K \geq 1$,

$$\Phi\left(0, yQ^N\right) - \mathbb{E}\left(\Phi\left(0, yQ^N\right)\right) = \sum_{p=0}^{Q^N-1}\left[\Phi(yp, yp+y) - \mathbb{E}(\Phi(0,y))\right] + \sum_{n=1}^{K}\sum_{k=0}^{Q^{N-n}-1}\left[\phi_{\mathcal{N}}\left(ykQ^n, y(k+1)Q^n\right) - \mathbb{E}\left(\phi_{\mathcal{N}}\left(0, yQ^n\right)\right)\right] + \Delta_K.$$

where Δ_K is the residual term of the series. By using the method to compute the variance, it is not difficult to establish that, for any $\varepsilon > 0$ there exists some $K > 0$ the expected value of $(\Delta_K/Q^N)^2$ is less that ε, for N sufficiently large.

By regrouping the terms of the above equation according to the Q-adic intervals $[yk/Q^K, y(k+1)/Q^K]$ for $0 \leq k < Q^{N-K}$ and by using the independence properties of the Poisson process $\mathcal{N}$, the quantity $\Phi\left(0, yQ^N\right) - \mathbb{E}\left(\Phi\left(0, yQ^N\right)\right) - \Delta_K$ can be written as a sum of Q^{N-K} independent identically distributed random variables. Therefore, the *classical* central limit theorem can be applied. The proposition is proved. $\square$

4.3. **The Distribution of the Sequence $(\mathbf{R_n})$.** The following proposition describes the distribution of the variableR_n in terms of n i.i.d. uniformly distributed random variables on the interval $[0,1]$. This characterization is generally implicitly used to get various asymptotics describing the depth of the associated tree. See Mahmoud [27] and Pittel [34].

Proposition 18. *For $n \geq 0$, the random variable R_n has the same distribution as*

$$R_n \overset{dist.}{=} 1 + Q \sum_{p \geq 0} \sum_{k=0}^{Q^p - 1} \phi_{\mathcal{U}_n}\left(]k/Q^p, (k+1)/Q^p]\right), \tag{30}$$

where, for $0 \leq a \leq b \leq 1$, $\phi_{\mathcal{U}_n}(]a,b]) = 1$ if $\mathcal{U}_n(]a,b]) \geq D$ and 0 otherwise. The variable $\mathcal{U}_n$ is the point measure on $[0,1]$ defined by

$$\mathcal{U}_n = \delta_{U_1} + \delta_{U_2} + \cdots + \delta_{U_n},$$

$(U_1, \ldots, U_n)$ are i.i.d. random variables uniformly distributed on $[0,1]$, in particular, $\mathcal{U}_n(]a,b])$ is the number of U_i's in the interval $]a,b]$.

Proof. Assume that $\mathcal{N}$ is a Poisson process with parameter 1, by definition

$$\left(R_{\mathcal{N}(]0,x])} \mid \mathcal{N}(]0,x]) = n\right) \overset{\text{dist.}}{=} R_n.$$

Due to Proposition 13, the distribution of the Poisson transform $R_{\mathcal{N}(]0,x])}$ is expressed as a functional of the points of the Poisson process on the interval $[0,x]$. But, as in the proof of Proposition 6, conditionally on the event $\{\mathcal{N}(]0,x]) = n\}$, these points can be expressed as xU_i, $1 \leq i \leq n$ where (U_i) are i.i.d. uniformly distributed random variables on the interval $[0,1]$. Equation (30) is thus a direct consequence of Relation (21) □

5. Some Results from Renewal Theory

Some definitions and results from renewal theory are briefly recalled. See Grimmett and Stirzaker [13] for example.

Definition 19. Arithmetic Distributions. *A distribution μ on $\mathbb{R}_+$ is* arithmetic *if there exists some $\lambda > 0$, such that $\mu\left(\{n\lambda : n \in \mathbb{N}\}\right) = 1$. The largest λ with this property is the* span *of the distribution.*

5.1. **Continuous Renewal Processes.** If (τ_i) is an i.i.d. sequence, for $x > 0$, the variable ν_x is defined as the hitting time of the set $[x, +\infty[$ by the random walk associated to (τ_i),

$$\nu_x = \inf\{n : S_n \geq x\},$$

with, for $n \geq 0$, $S_n = \tau_1 + \tau_2 + \cdots + \tau_n$.

The variable $t_0^x = S_{\nu_x} - x$ is the overshoot of level x by the random walk. For $k \geq 0$, one defines t_k^x as the location of the kth point before x, $t_k^x = S_{\nu_x - k} - x$, with the convention that $S_i = -\infty$ for $i < 0$.

Theorem 20 (Continuous Renewal Theorem)**.** *If the distribution of τ_0 is non-arithmetic, then the random variables*

$$\left((t_k^x - t_{k+1}^x, k \geq 1), t_1^x, t_0^x\right)$$

converges in distribution to a sequence $((\tilde{\tau}_i, i \geq 2), \tau_1^*, \tau^*)$, *where the sequence* $(\tilde{\tau}_i, i \geq 2)$ *is i.i.d. with the same distribution as* τ_0 *and independent of the variables* (τ_1^*, τ^*) *whose distribution is given by*

$$\mathbb{E}\left(f(\tau_1^*, \tau^*)\right) = \frac{1}{\mathbb{E}(\tau_0)} \mathbb{E}\left(\int_0^{\tau_0} f(-u, \tau_0 - u)\, du\right), \tag{31}$$

for any non-negative Borelian function f *on* $\mathbb{R}^2$.

5.2. **Discrete Renewal Processes.** If (C_i) is an i.i.d. sequence of integer valued random variables such that $\mathbb{P}(C_1 = 1) > 0$ and

$$\tau_n = \inf\left\{k : \sum_{i=1}^{k} C_i \geq n\right\},$$

Theorem 21 (Discrete Renewal Theorem)**.** *The random variables*

$$\left((C_{\tau_n - k}, 1 \leq k \leq \tau_n), n - \sum_{i=1}^{\tau_n - 1} C_i\right)$$

converges in distribution to $((\tilde{C}_i, i \geq 2), C_1^*)$, *where the sequence* $(\tilde{C}_i, i \geq 2)$ *is i.i.d. with the same distribution as* C_1 *and independent of the variable* C_1^* *whose distribution is given by*

$$\mathbb{P}(C_1^* = n) = \frac{1}{\mathbb{E}(C_1)} \mathbb{P}(C_1 \geq n), \qquad n \geq 1.$$

References

[1] Thomas H. Cormen and Charles E. Leiserson and Ronald L. Rivest, *Algorithms*, MIT Press, Cambridge, Massachusetts, 1990.

[2] Julien Barral, *Moments, continuité et analyse multifractale des martingales de mandelbrot*, Probability Theory and Related Fields **113** (1999), no. 4, 582–597.

[3] Jean Bertoin, *Homogeneous fragmentation processes*, Probability Theory and Related Fields **121** (2001), 301–318.

[4] John I. Capetanakis, *Tree algorithms for packet broadcast channels*, IEEE Transactions on Information Theory **25** (1979), no. 5, 505–515.

[5] Joseph T. Chang, *Inequalities for the overshoot*, The Annals of Applied Probability **4** (1994), no. 4, 1223–1233.

[6] J. Clément, P. Flajolet, and B. Vallée, *Dynamical sources in information theory: a general analysis of trie structures*, Algorithmica **29** (2001), no. 1-2, 307–369, Average-case analysis of algorithms (Princeton, NJ, 1998).

[7] Luc Devroye, *Universal limit laws for depths in random trees*, SIAM Journal on Computing **28** (1999), no. 2, 409–432 (electronic).

[8] Luc Devroye, *Universal asymptotics for random tries and Patricia trees*, preprint, March 2004.

[9] Anthony Ephremides and Bruce Hajek, *Information theory and communication networks: an unconsummated union*, IEEE Transactions on Information Theory **44** (1998), no. 6, 1–20.

[10] Kenneth Falconer, *Techniques in fractal geometry*, John Wiley and Sons, 1997.

[11] G. Fayolle, P. Flajolet, and M. Hofri, *On a functional equation arising in the analysis of a protocol for a multi-access broadcast channel*, Advances in Applied Probability **18** (1986), 441–472.

[12] Philippe Flajolet, Xavier Gourdon, and Philippe Dumas, *Mellin transforms and asymptotics: harmonic sums*, Theoretical Computer Science **144** (1995), no. 1-2, 3–58, Special volume on mathematical analysis of algorithms.

[13] G. R. Grimmett and D. R. Stirzaker, *Probability and random processes*, second ed., The Clarendon Press Oxford University Press, New York, 1992.

[14] B.M. Hambly and Michel L. Lapidus, *Random fractal strings: their Zeta functions, complex dimensions and spectral asymptotics*, Transactions of the AMS (2004), –.
[15] Micha Hofri, *Analysis of algorithms*, The Clarendon Press Oxford University Press, New York, 1995, Computational methods & mathematical tools.
[16] Philippe Jacquet and Mireille Régnier, *Trie partitioning process: Limiting distributions*, Lecture Notes in Computer Science, vol. 214, Springer Verlag, New York, 1986, pp. 196–210.
[17] Svante Janson and Wojciech Szpankowski, *Analysis of an asymmetric leader election algorithm*, Electronic Journal of Combinatorics **4** (1997), no. 1, Research Paper 17, 16 pp. (electronic).
[18] J.P. Kahane and J. Peyrière, *Sur certaines martingales de benoît mandelbrot*, Advances in Mathematics **22** (1976), 131–145.
[19] J. F. C. Kingman, *Poisson processes*, Oxford Studies in Probability, 1993.
[20] Peter Kirschenhofer, Helmut Prodinger, and Wojciech Szpankowski, *Analysis of a splitting process arising in probabilistic counting and other related algorithms*, Random Structures and Algorithms **9** (1996), no. 4, 379–401.
[21] Charles Knessl and Wojciech Szpankowski, *Limit laws for the height in PATRICIA tries*, Journal of Algorithms **44** (2002), no. 1, 63–97, Analysis of algorithms.
[22] Donald E. Knuth, *The art of computer programming. Volume 3*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973, Sorting and searching, Addison-Wesley Series in Computer Science and Information Processing.
[23] M.L. Lapidus and M. van Frankenhuysen, *Fractal geometry and number theory*, Birkhaüser, Boston, 2000.
[24] Q. Liu, *On generalized multiplicative cascades*, Stochastic Processes and their Applications **86** (2000), 263–286.
[25] Gary Lorden, *On excess over the boundary*, Annals of Mathematical Statistics **41** (1970), 520–527.
[26] Guy Louchard and Helmut Prodinger, *The moments problem of extreme-value related distribution functions*, 2004, Preprint.
[27] Hosam M. Mahmoud, *Evolution of random search trees*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., New York, 1992, A Wiley-Interscience Publication.
[28] Benoît Mandelbrot, *Multiplications aléatoires et distributions invariantes par moyennes pondérées*, Comptes Rendus de l'Académie des Sciences Série I **278** (1974), 289–292 and 355–358.
[29] P. Mathys and P. Flajolet, *Q-ary collision resolution algorithms in random access systems with free or blocked channel access*, IEEE Transactions on Information Theory **31** (1985), 244–254, Special Issue on Random Access Communication.
[30] R. Daniel Mauldin and S.C. Williams, *Random recursive constructions: Asymptotic geometric and topological properties*, Transactions of the AMS **295** (1986), no. 1, 325–346.
[31] Grégory Miermont, *Coalescence et fragmentation stochastique, arbres aléatoires et processus de Lévy*, Ph.D. thesis, Université de Paris VI, December 2003.
[32] Hanène Mohamed and Philippe Robert, *Analysis of dynamic tree algorithms*, 2004, In preparation.
[33] Jim Pitman, *Exchangeable and partially exchangeable random partitions*, Probability Theory and Related Fields **102** (1995), 145–158.
[34] B. Pittel, *Asymptotical growth of a class of random trees*, Annals of Probability **13** (1985), no. 2, 414–427.
[35] Danny Raz, Yuval Shavitt, and Lixia Zhang, *Distributed council election*, IEEE/ACM Transactions on Networking (2004), –.
[36] Philippe Robert, *On the asymptotic behavior of some algorithms*, Random Structures and Algorithms **xx** (2004), no. xx, xx, To Appear.
[37] Mireille Régnier and Philippe Jacquet, *New results on the size of tries*, IEEE Transactions on Information Theory **35** (1989), 203–205.
[38] Robert Sedgewick and Philippe Flajolet, *Introduction to the analysis of algorithms*, Addison-Wesley, 1995.
[39] Wojciech Szpankowski, *Average case analysis of algorithms on sequences*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2001.

[40] B. S. Tsybakov and V. A. Mikhaĭlov, *Free synchronous packet access in a broadcast channel with feedback*, Problems Inform. Transmission **14** (1978), no. 4, 32–59.

[41] Edward C. Waymire and S.C. Williams, *A general decomposition theory for random cascades*, Bulletin of the AMS **31** (1994), no. 2, 216–222.

[42] Jack K. Wolf, *Born again group testing: multiaccess communications*, IEEE Transactions on Information Theory **31** (1985), no. 2, 185–191.

E-mail address: `Hanene.Mohamed@inria.fr`

E-mail address: `Philippe.Robert@inria.fr`
URL: `http://www-rocq.inria.fr/~robert`

(H. Mohamed, Ph. Robert) INRIA, domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex, France